To appear in the *International Journal of Control*
Vol. 00, No. 00, Month 20XX, 1–17

RESEARCH ARTICLE

A Homotopy-based Moving Horizon Estimation

Mohammad Abdollahpouri[a*], Rien Quirynen[c†], Mark Haring[b], Tor Arne Johansen[b], Gergely Takács[a], Moritz Diehl[c] and Boris Rohaľ-Ilkiv [a]

[a]*Institute of Automation, Measurement and Applied Informatics, Faculty of Mechanical Engineering, Slovak University of Technology in Bratislava, 812 31 Bratislava, Slovakia;*
[b]*Center for Autonomous Marine Operations and Systems (AMOS), Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7491 Trondheim, Norway;*
[c]*Department of Microsystems Engineering IMTEK, University of Freiburg, 79110 Freiburg, Germany.*

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

Moving horizon estimation (MHE) solves a constrained dynamic optimization problem at each sampling instant. Including nonlinear dynamics into an optimal estimation problem generally comes at the cost of tackling a non-convex optimization problem. In this article, a particular model formulation is proposed in order to convexify a class of nonlinear MHE problems. It delivers a linear time varying (LTV) model that is globally equivalent to the nonlinear dynamics in a noise-free environment, hence the optimization problem becomes convex, where its solution is the global minimum. On the other hand, in the presence of noise and unknown disturbances, the accuracy of the LTV model degrades and this results in a less accurate solution for the original nonlinear constrained optimization problem. This approximation may drive the convex solution away from the global minimum in the case of unknown disturbances. For this purpose, considering some assumptions on these bounded disturbances, a homotopy based approach is proposed in order to transform the problem from the convex to the non-convex problem formulation, where the sequential implementation of this technique starts with solving the convexified MHE problem. Each step in the proposed homotopy procedure requires the solution of a dynamic optimization problem, which can be used as an initial guess for the next problem formulation. This procedure continues until a locally optimal solution to the original non-convex estimation problem is found. The first simulation study validates the efficiency of the proposed approach in obtaining a solution close to the global minimum with zero mean random disturbances. In addition, the second case study illustrates that in the presence of large biased disturbances, the result of the convexified MHE could lead the homotopy based MHE to an alternative locally optimal solution.

Keywords: moving horizon estimation, non-convex optimization, nonlinear systems, nonlinear state and parameter estimation.

1. Introduction

One of the main challenges in nonlinear state and parameter estimation is the non-convex nature of the dynamic optimization problem, which could result in a locally optimal solution. For instance, the estimates of the well-known extended Kalman filter (EKF) can converge to a wrong value with a poor initial guess; see the case studies by Johansen & Fossen (2016a); Haseltine & Rawlings (2005). A convexified approximation is one of the approaches that has been considered in many works to tackle a non-convex problem; e.g. see the works by Diehl et al. (2009); Dinh et al. (2011); Hovgaard et al. (2013).

*Mohammad Abdollahpouri is currently affiliated with the Chalmers University of Technology in Göteborg, Sweden. Email: mohammad.abdollahpouri@chalmers.se
†Rien Quirynen is currently affiliated with the Mitsubishi Electric Research Laboratories (MERL) in Cambridge, MA.

Moving horizon estimation (MHE), as one of the candidates for online constrained state estimation, can deal with nonlinear dynamics explicitly and directly incorporate the physical or logical constraints in its design. The MHE problem formulation is an approximation of a full information constrained estimation technique. It considers a fixed number of measurements inside a moving time window and this makes it tractable in practice, which is almost impossible for a full information estimation approach. The *arrival cost* term is typically introduced in the MHE formulation to represent the truncated data; see the explanation by Rao et al. (2003) for more details on MHE. Its increased performance compared to classical techniques, such as Kalman filtering, has been studied extensively on several case studies; e.g. see the results on chemical engineering systems by Haseltine & Rawlings (2005) and on vibration systems by Abdollahpouri et al. (2016); Abdollahpouri et al. (2017). There are several factors that make MHE quite practical in industry compared to the traditional nonlinear estimation techniques (Johansen, 2011): including N samples of measurements in the moving time window and incorporating the physical and logical constraints in a dynamic optimization framework. One of the challenges that needs to be addressed from an optimization point of view, is to find a good local solution (depending on the case study) or ultimately to obtain a global solution to a non-convex problem reliably. Most available algorithms for treating non-convex optimization problems are gradient based and end up with a local minimum. Sub-optimal solutions not only degrade the quality of estimates, they can be misleading as well, e.g., in a structural health monitoring system, by utilizing the incorrect state and parameters estimates.

A model reformulation has been studied by Johansen & Fossen (2016a), where the nonlinear dynamics are translated into a linear time varying (LTV) model without any local linearization of the dynamics. This technique is conceptually similar to the differential flatness theory; see the works presented by Fliess et al. (1995) and Sira-Ramirez & Agrawal (2004). This approach eliminates the nonlinearity by utilizing a set of measurements without optimally considering the input and output disturbances. The resulting LTV model, in its nominal formulation, is globally equivalent to the nonlinear system. Using this model transformation, Johansen & Fossen (2016b) proposed a two-stage Kalman filtering method, which is called a double Kalman filter (DKF). This approach has been analyzed for continuous time and recently its stability analysis in discrete time has been studied in Abdollahpouri et al. (2017). The cascade configuration in DKF utilizes the transformed LTV model for sub-optimal estimation in the first stage, which then provides a linearization point for the second stage. In the second stage, a linearized Kalman filter (LKF) further improves the estimation quality that has been lost because of the sub-optimal consideration of input and output disturbances in the first stage.

One approach for tackling a non-convex optimization problem for receding horizon strategies has been proposed by Bonilla et al. (2010b). This work assumed no input or output disturbances, a dynamical model that is affine in the parameters and full state measurements, which allows one to transform the nonlinear dynamics to become linear in the parameters. Therefore, considering this linear model instead, the resulting optimization problem becomes convex. The solution of this convex approximation can be used as an initialization for the original non-convex problem. This results in a homotopy based automatic initialization technique for parameter estimation problems (Bonilla et al., 2010a), when all the states are assumed to be measured. It is worth mentioning that our approach does not require all states to be measured and is therefore more generally applicable.

Further on this topic, Liu (2013) proposed an MHE scheme that makes use of a deterministic observer for nonlinear systems with bounded model uncertainties. The auxiliary nonlinear observer is meant to provide additional information to the arrival cost approximation and hence improve the state estimates. However, the auxiliary nonlinear observer is locally convergent and its sensitivity to different initializations may provide wrong information to the MHE scheme, thus the overall approach has some local convergence properties that can be inherited from the auxiliary observer; readers are encouraged to compare it with the DKF (Johansen & Fossen, 2016b). A more detailed comparison in practice, between the proposed approach and other MHE formulations, such as the one in Liu (2013), is part of ongoing research.

In this paper, we propose a constrained nonlinear estimation approach, which is globally convergent in the disturbance-free case based on the reformulation technique by Johansen & Fossen (2016b). This flatness-like model transformation is utilized to globally transform the nonlinear dynamics into an LTV model. The reformulation is applied to MHE, and results in a convexified optimization problem (CMHE) that, however, becomes suboptimal in case of input and output disturbances. By the use of a homotopy approach, the suboptimal solution from CMHE can be gradually improved by solving a sequence of dynamic optimization problems, resulting in an optimal solution for the original non-convex estimation problem. Even with the added complexity of the homotopy based MHE (HMHE), the computational burden of our method is comparable or can even become smaller than in case the nonlinear optimization problems are solved in a cold started fashion. This is because the proposed method starts by solving the CMHE and uses the result for initializing the next steps. First, let us begin by introducing the problem formulation.

2. Flatness based problem formulation

This section provides the definitions and assumptions, needed for employing the model transformation. The nonlinear dynamics are described by

$$x_{k+1} = f(x_k) + \acute{w}_k, \tag{1a}$$
$$y_k = h(x_k) + \acute{v}_k, \tag{1b}$$

where $x_k \in \mathbb{X} \subseteq \mathbb{R}^n$, $\acute{w}_k \in \mathbb{W} \subseteq \mathbb{R}^n$ and $\acute{v}_k \in \mathbb{V} \subseteq \mathbb{R}^p$ respectively denote the state vector, input and output disturbances in discrete time. The nonlinear dynamics are defined by $f(\cdot) : \mathbb{X} \to \mathbb{X}$ and the nonlinear measurement function is expressed by $h(\cdot) : \mathbb{X} \to \mathbb{R}^p$. Furthermore, $y_k \in \mathbb{R}^p$ is the bounded observed output and the discrete time index is denoted by k. The solution for the dynamic system in (1) is denoted by $x(k; x_0, 0, (\acute{w}_j)_{j=0}^{k-1})$ at time k, starting from time point 0 with initial value x_0 and given the sequence of input disturbances $(\acute{w}_j)_{j=0}^{k-1}$. The output response is denoted by $y(k; x_0, 0, (\acute{w}_j)_{j=0}^{k-1}) := h(x(k; x_0, 0, (\acute{w}_j)_{j=0}^{k-1}))$, such that $\acute{v}_k := y_k - y(k; x_0, 0, (\acute{w}_j)_{j=0}^{k-1})$. For a disturbance-free case, $(\acute{w}_j)_{j=0}^{k-1} = 0$, the nominal output response is denoted by $y(k; x_0, 0)$.

Assumption 1: *The sets $\mathbb{W}$ and $\mathbb{V}$ are compact with $0 \in \mathbb{W}$ and $0 \in \mathbb{V}$.*

Assumption 2: *For any possible sequence of disturbances $(\acute{w}_j)_{j=0}^{k-1}$, the initial state x_0 leads to the system trajectory $x(k; x_0, 0, (\acute{w}_j)_{j=0}^{k-1})$ that lies in a compact set $\mathbb{X}$.*

Assumption 3: *The nonlinear functions $f(\cdot)$ and $h(\cdot)$ are Lipschitz continuous in all of their arguments with Lipschitz constants c_f and c_h, respectively, for all $k \geq 0$.*

Assumption 4: *There exists a map $\psi(\cdot, \cdot, \cdot) : \mathbb{R}^{pd} \times \mathbb{R}^{n(d-1)} \times \mathbb{R}^{pd} \to \mathbb{R}^n$, which is Lipschitz continuous in all of its arguments, and there exists a positive integer d such that the state of (1) for all $k \geq d-1$ can be written as*

$$x_k = \psi\big((y_j)_{j=l}^{k}, (\acute{w}_j)_{j=l}^{k-1}, (\acute{v}_j)_{j=l}^{k}\big),$$

where $l = k - d + 1$.

Remark 1: Assumption 4 implies that system (1) is *difference flat* (Sira-Ramirez & Castro-Linares, 2000), with flat outputs $(y_j)_{j=l}^{k}$, $(\acute{w}_j)_{j=l}^{k-1}$ and $(\acute{v}_j)_{j=l}^{k}$. Moreover, it follows from Assumption 4 that the current state x_k can be uniquely determined from a number of current and past output measurements $(y_j)_{j=l}^{k}$ and disturbances $(\acute{w}_j)_{j=l}^{k-1}$ and $(\acute{v}_j)_{j=l}^{k}$. This is naturally related to the observability

of the dynamic system in (1).

Definition 1: (Rao et al., 2003) A function $\phi : \mathbb{R} \to \mathbb{R}$ is a K-function, if it is continuous, strictly monotone increasing, positive definite and unbounded.

Definition 2: (Rao et al., 2003) The system in (1) is said to be uniformly observable if there exists a positive integer N and a K-function $\phi(\cdot)$ for any $x_1, x_2 \in \mathbb{X}$, with

$$\phi(\|x_1 - x_2\|) \leq \sum_{j=0}^{N-1} \|y(k+j; x_1, k) - y(k+j; x_2, k)\|,$$

for all $k \geq 0$.

3. The proposed model transformation

Assuming the nonlinear system (1) satisfies Assumption 3 and 4, i.e. difference flatness and s-moothness properties, this section provides the information needed to perform the LTV model transformation. Let us use Proposition 2.4.7 in Abraham et al. (2012) for (1)

$$f(x_k) = f(0) + \int_0^1 \frac{\partial f}{\partial x}(sx_k)ds\, x_k = f(0) + \int_0^1 \frac{\partial f}{\partial x}\Big(s\psi\big((y_j)_{j=l}^{k}, (\acute{w}_j)_{j=l}^{k-1}, (\acute{v}_j)_{j=l}^{k}\big)\Big)ds\, x_k,$$
$$h(x_k) = h(0) + \int_0^1 \frac{\partial h}{\partial x}(sx_k)ds\, x_k = h(0) + \int_0^1 \frac{\partial h}{\partial x}(s\psi((y_j)_{j=l}^{k}, (\acute{w}_j)_{j=l}^{k-1}, (\acute{v}_j)_{j=l}^{k}))ds\, x_k,$$

using Assumption 4, in order to replace the nonlinear terms in the state space model by the associated measurements. When there are no disturbances, $x_k = \psi((y_j)_{j=l}^{k}, 0, 0)$, then the nonlinear dynamics can be globally transformed to an LTV model as

$$x_{k+1} = f(0) + F\big((y_j)_{j=l}^{k}\big)x_k + w_k, \tag{2a}$$
$$y_k = h(0) + H\big((y_j)_{j=l}^{k}\big)x_k + v_k, \tag{2b}$$

where $F_k = F(\cdot)$ and $H_k = H(\cdot)$ are time varying functions of $(y_j)_{j=l}^{k}$. It should be noted that the state vector does not change by applying the linearizing transformation; hence the same notation is utilized. One possible approach for finding F_k and H_k is given as

$$F_k = \int_0^1 \frac{\partial f}{\partial x}\Big(s\psi\big((y_j)_{j=l}^{k}, 0, 0\big)\Big)ds, \quad H_k = \int_0^1 \frac{\partial h}{\partial x}\Big(s\psi\big((y_j)_{j=l}^{k}, 0, 0\big)\Big)ds,$$

although simpler, but problem specific methods may be possible in many practical applications. Note that after this transformation, the input and output disturbances are accounted for by w_k and v_k, respectively. Although transforming (1) to (2) keeps the nominal dynamics unaffected, the input and measurement disturbances are not entering the model (2) in the same way as in (1); $w_k = w\big((y_j)_{j=l}^{k}, (\acute{w}_j)_{j=l}^{k-1}, (\acute{v}_j)_{j=l}^{k}\big)$ and $v_k = v\big((y_j)_{j=l}^{k}, (\acute{w}_j)_{j=l}^{k-1}, (\acute{v}_j)_{j=l}^{k}\big)$. This can be summarized as

follows

$$f(0)+F_kx_k+w_k=f(x_k)+\acute{w}_k\rightarrow$$
$$w_k=\acute{w}_k+\int_0^1\Big(\frac{\partial f}{\partial x}(s\psi\big((y_j)_{j=l}^k,(\acute{w}_j)_{j=l}^{k-1},(\acute{v}_j)_{j=l}^k\big))-\frac{\partial f}{\partial x}(s\psi((y_j)_{j=l}^k,0,0))\Big)ds\ x_k, \quad (3)$$
$$h(0)+H_kx_k+v_k=h(x_k)+\acute{v}_k\rightarrow$$
$$v_k=\acute{v}_k+\int_0^1\Big(\frac{\partial h}{\partial x}(s\psi\big((y_j)_{j=l}^k,(\acute{w}_j)_{j=l}^{k-1},(\acute{v}_j)_{j=l}^k\big))-\frac{\partial h}{\partial x}(s\psi((y_j)_{j=l}^k,0,0))\Big)ds\ x_k, \quad (4)$$

where we note that $\int_0^1\frac{\partial f}{\partial x}(sx_k)ds\ x_k=f(x_k)-f(0)$ and $\int_0^1\frac{\partial h}{\partial x}(sx_k)ds\ x_k=h(x_k)-h(0)$, which follows from Abraham et al. (2012, Prop. 2.4.7). From Assumption 4, we have that $x_k=\psi((y_j)_{j=l}^k,0,0)$ if $\acute{w}_k$ and $\acute{v}_k$ are zero for all k. By substituting $x_k=\psi((y_j)_{j=l}^k,0,0)$ in (3) and (4), one can directly see that w_k and v_k are uniformly zero if $\acute{w}_k$ and $\acute{v}_k$ are uniformly zero. Remark 2 quantifies the deviation between the state obtained by the approximated LTV model from the true state value, based on the input and output disturbance bounds.

Remark 2: If the elements of $(y_j)_{j=l}^k$, $(w_j)_{j=l}^{k-1}$ and $(v_j)_{j=l}^k$ are uniformly bounded $\|(w_j)_{j=l}^{k-1}\|\leq C_w$, $\|(v_j)_{j=l}^k\|\leq C_v$ and ψ is Lipschitz continuous with respect to $(w_j)_{j=l}^{k-1}$ and $(v_j)_{j=l}^k$ with the constant L_ψ, then the corresponding estimation error is bounded by

$$\|x_k-\psi\big((y_j)_{j=l}^k,0,0\big)\|\leq L_\psi\big(\|(w_j)_{j=l}^{k-1}\|+\|(v_j)_{j=l}^k\|\big):=C_\psi,$$

where $C_\psi=L_\psi(C_w+C_v)$.

Assumption 5: *The LTV model in* (2) *is observable in the sense of Definition 2.*

Example 1: Let the nonlinear dynamics of (1) be described in discrete time as

$$\begin{aligned}x_{k+1,1}&=x_{k,1}+x_{k,2}+w_{k,1},\\ x_{k+1,2}&=-x_{k,1}x_{k,2}^2+x_{k,1}\sin^2(x_{k,1}),\\ y_k&=x_{k,1}+v_k.\end{aligned}$$

Let $\delta_k=y_k-v_k$ such that $x_{k,1}=\delta_k$. From the first system equation, we obtain $x_{k-1,2}=\delta_k-\delta_{k-1}-w_{k-1,1}$. Substituting this in the second system equation yields $x_{k,2}=-\delta_{k-1}(\delta_k-\delta_{k-1}-w_{k-1,1})^2+\delta_{k-1}\sin^2(\delta_{k-1})$. Hence, we can define ψ in Assumption 4 as

$$\begin{aligned}\psi\big((y_j)_{j=k-1}^k,(\acute{w}_j)_{j=k-1}^{k-1},(\acute{v}_j)_{j=k-1}^k\big)&=\begin{bmatrix}\delta_k\\ -\delta_{k-1}(\delta_k-\delta_{k-1}-w_{k-1})^2+\delta_{k-1}\sin^2(\delta_{k-1})\end{bmatrix}\\ &:=\begin{bmatrix}\phi_1(y_k,v_k)\\ \phi_2\big((y_j)_{j=k-1}^k,(\acute{w}_j)_{j=k-1}^{k-1},(\acute{v}_j)_{j=k-1}^k\big)\end{bmatrix}.\end{aligned}$$

Using the map ψ, the nonlinear system can be written as an LTV system of the form (2), with $f(0)=0$, $h(0)=0$, $\phi_{k,1}:=\phi_1(y_k,0)$ and $\phi_{k,2}:=\phi_2(y_k,0,0)$

$$\begin{aligned}x_{k+1,1}&=x_{k,1}+x_{k,2}+w_{1,k},\\ x_{k+1,2}&=x_{k,1}\sin^2(\phi_{1,k})-\phi_{1,k}\phi_{2,k}x_{k,2}+w_{2,k},\\ y_k&=x_{k,1}+v_k,\end{aligned}$$

as in (2a) and (2b). We note that this linearizing transformation is not unique, since the second state update can be rewritten, e.g., as $x_{k+1,2} = (-(\eta x_{k,2}^2) + \sin^2(x_{k,1}))x_{k,1} + (1-\eta)(-x_{k,1}x_{k,2})x_{k,2}$ for any scalar η. Hence, the use of any matrix F_k that satisfies

$$F_k = \begin{bmatrix} 1 & 1 \\ \sin^2(\phi_{1,k}) - \eta\phi_{2,k}^2 & -(1-\eta)\phi_{1,k}\phi_{2,k} \end{bmatrix},$$

for any $\eta \in \mathbb{R}$, results in an LTV system. □

4. Homotopy based Moving Horizon Estimation

When using the LTV dynamic model (2) for estimation, one obtains a convex optimization problem with a unique global minimum. The solution obtained from this problem will be used as a starting point for the nonlinear MHE formulation. The proposed approach is based on several intermediate steps between the solution of the convex problem and the non-convex MHE formulation, in order to perform a numerical continuation on the resulting solution manifold (Allgower & Georg, 1990). In what follows, we will specify the nonlinear MHE formulation, which generally is a non-convex optimization problem. Furthermore, using the transformed LTV model, the convexified problem can be formulated and we present the corresponding homotopy based approach.

4.1 *Nonlinear MHE formulation*

First, we formulate the nonlinear moving horizon estimation as the following optimal control problem (OCP) in discrete time by defining $l = k - N + 1$

$$\begin{aligned} \text{MHE}: \quad \Phi_k = \min_{(\hat{x}_j)_{j=l}^k, (\acute{\xi}_j)_{j=l}^{k-1}, (\acute{\nu}_j)_{j=l}^k} \; & \Gamma_l(\hat{x}_l) + \sum_{i=l}^{k} L_i(\acute{\xi}_i, \acute{\nu}_i) \\ \text{s.t.} \quad & \hat{x}_{i+1} = f(\hat{x}_i) + \acute{\xi}_i, && i = l, \dots, k-1, && \text{(5a)} \\ & y_i = h(\hat{x}_i) + \acute{\nu}_i, && i = l, \dots, k, && \text{(5b)} \\ & \left((\hat{x}_j)_{j=l}^k, (\acute{\xi}_j)_{j=l}^{k-1}\right) \in \Omega \end{aligned}$$

where $\hat{x}_k$ denotes the current estimated state of the MHE problem formulation, Ω is the convex set of constraints defined in the bottom of this page[‡]. The stage cost is denoted by $L_i(\cdot,\cdot) \in \mathbb{R}$, and the arrival cost $\Gamma_l(\cdot) \in \mathbb{R}$ can be calculated based on an update procedure as found in the literature, e.g., see a linearization based approach by Kühl et al. (2011). In Fig. 1, the complete scheme of the MHE implementation is illustrated, where the measurements $(y_j)_{j=l}^k$ are stored using an *N-step data buffer* block. Furthermore, the first element of the measurement set y_l is utilized by the *arrival cost update* block to provide the *optimizer* with the necessary parameters; see Kühl et al. (2011). Finally, $(x_i)_{i=l}^k$ denotes the optimized trajectory of state values over the current time horizon.

Assumption 6: *The arrival cost term is assumed to be a positive definite quadratic function as* $\Gamma_l(\hat{x}_l) := \|\hat{x}_l - \tilde{x}_l\|_{P_l}^2$, *where* $\tilde{x}_l$ *and* $P_l \succ 0$ *are the online tuning parameters.*

Assumption 7: *Let the stage cost* $L_i(\acute{\xi}_i, \acute{\nu}_i) := \acute{\xi}_i^{\mathrm{T}} \acute{Q}_i^{-1} \acute{\xi}_i + \acute{\nu}_i^{\mathrm{T}} \acute{R}_i^{-1} \acute{\nu}_i$ *for* $i = l, \dots, k-1$ *and* $L_k(\acute{\xi}_k, \acute{\nu}_k) = \acute{\nu}_k^{\mathrm{T}} \acute{R}_k^{-1} \acute{\nu}_k$, *where* $\acute{Q}_i \succ 0$ *and* $\acute{R}_i \succ 0$ *are the tuning parameters with proper dimensions.*

[‡] $\Omega := \left\{ (x, \xi) : \begin{array}{lcll} x_i & \in & \mathbb{X}, & i = l, \dots, k \\ \xi_i & \in & \mathbb{W}, & i = l, \dots, k-1 \\ \nu_i & \in & \mathbb{V}, & i = l, \dots, k \end{array} \right.$

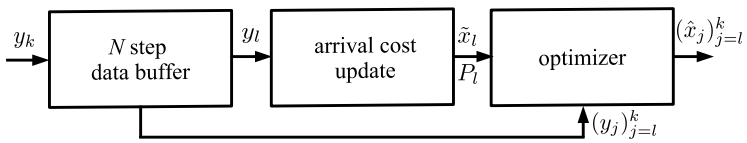


Figure 1. Schematic overview of nominal MHE.

Assumption 8: *There exists a $\mathcal{K}$-function γ such that*

$$0 \leq \Gamma_k(x) \leq \gamma(\|x - \hat{x}_k\|)$$

holds for all $x \in \mathbb{X}$.

Considering the reachable set $\mathcal{R}_k^N = \{x(k; x_l, l, (\acute{w}_j)_{j=l}^{k-1}) : ((x_j)_{j=l}^k, (\acute{w}_j)_{j=l}^{k-1} \in \Omega)\}$ at time k with a feasible initial state (over the horizon) x_l and an input disturbance sequence of $(w_j)_{j=l}^{k-1}$, the following assumption is complementary for a bounded estimation error of the MHE.

Assumption 9: *For any $k \geq N$, and any $p \in \mathcal{R}_k^N$, the arrival cost $\Gamma_k(\cdot)$ satisfies the following condition*

$$\Gamma_k(p) \leq \min_{(\hat{x}_j)_{j=l}^k, (\acute{\xi}_j)_{j=l}^{k-1}, (\acute{\nu}_j)_{j=l}^k} \left\{ \Gamma_l(\hat{x}_l) + \sum_{i=l}^{k} L_i(\acute{\xi}_i, \acute{\nu}_i) \ : x(k; x_l, l, (\acute{\xi}_j)_{j=l}^{k-1}) = p, \ \left((\hat{x}_j)_{j=l}^k, (\acute{\xi}_j)_{j=l}^{k-1}\right) \in \Omega \right\}$$

Lemma 1: *If Assumptions 1, 2, 3, 6 and 7 hold, and if the system in (1) is uniformly observable in the sense of Def. (2), and the arrival cost sequence satisfies the Assumptions 8 and 9, then the estimation error of MHE is bounded.*

Proof. The constraints are assumed to be convex, the state trajectory with arbitrary disturbances stays feasible and the nonlinear dynamics are sufficiently smooth. The stage cost function and arrival cost are assumed to be positive definite and quadratic. Therefore, all the assumptions are satisfied for applicability of Rao et al. (2003, Proposition 3.7), where its proof is dependent on the results of Rao et al. (2003, Proposition 3.4) and Rao et al. (2003, Lemma 2.5). □

4.2 *Convexified MHE formulation*

As a first step to deal with the non-convex optimization problem in (5), we could use the linear dynamic model from Eq. (2) instead of the nonlinear dynamics. This results in a convex MHE problem, if additionally the inequality constraints can be represented by convex sets (Boyd & Vandenberghe, 2004).

$$\text{CMHE}: \quad \min_{(\hat{x}_j)_{j=l}^k, (\xi_j)_{j=l}^{k-1}, (\nu_j)_{j=l}^k} \ \Gamma_l(\hat{x}_l) + \sum_{i=l}^{k} L_i(\xi_i, \nu_i) \tag{6a}$$

$$\text{s.t.} \quad \hat{x}_{i+1} = f(0) + F_i \hat{x}_i + \xi_i, \qquad i = l, \ldots, k-1, \tag{6b}$$

$$y_i = h(0) + H_i \hat{x}_i + \nu_i, \qquad i = l, \ldots, k, \tag{6c}$$

$$\left((\hat{x}_j)_{j=l}^k, (\xi_j)_{j=l}^{k-1}\right) \in \Omega, \tag{6d}$$

where $L_i(\cdot, \cdot)$ satisfies Assumption 7. It is assumed that ξ_j is limited in the same convex set as defined in the original MHE formulation of Eq. (5). Based on the explanation provided in Sect. 3, we know that the characteristics of w_j (analogous to ξ_j) are different from $\acute{w}_j$ (analogous to $\acute{\xi}_j$),

hence ξ_j and $\acute{\xi}_j$ not necessarily reside in the same set. However, in this method, we assume that they are in the same set and this can be a conservative assumption.

Remark 3: Given no input and output disturbances, the result of CMHE is equivalent to the unique solution of the nonlinear MHE problem. However, even with a small bounded disturbance, the two solutions of CMHE and MHE will be different. In Figure 2, x_1^C and x_2^C represent two possible solutions of CMHE with different input and output disturbances. Similarly, x_1^N and x_2^N are possible local and global minima of the nonlinear MHE problem formulation. This questions the guarantee in ensuring a global minimum, as Sect. 5.2 will illustrate this issue.

Remark 4: From Remark 3.2 in (Rao et al., 2003) the Assumption 9 is satisfied, since the arrival cost term is a positive quadratic function, and the system is described by an LTV model with convex constraints. Therefore, the boundedness analysis for CMHE estimates can be obtained similarly to Lemma 1 using Assumption 5 for the observability condition.

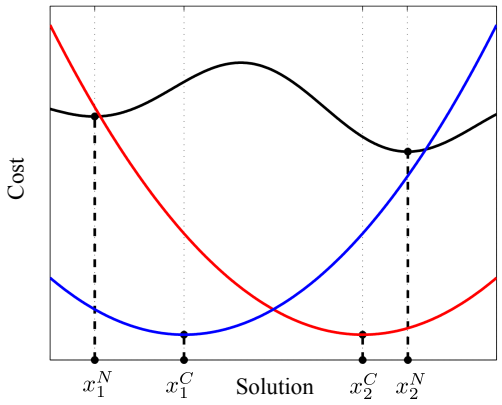


Figure 2. The illustration of local and global solutions for an arbitrary cost function; black line: scalar original (non-convex) cost function, blue line: scalar convexified cost function leading to the local minimum, red line: scalar convexified cost function leading to the global minimum.

4.3 *Homotopy based MHE implementation*

After introducing MHE and CMHE, we can propose an implementation based on the combination of those problems using a homotopy approach. This homotopy based approach—HMHE— is utilizing MHE and CMHE in one unified optimization problem. For this purpose, a tuning parameter λ has been introduced in (7a) and the homotopy based optimization problem is given as a sequence of

problems for $\lambda_j \in \Lambda$:

$$\text{HMMHE}(\lambda_j): \quad \min_{\substack{(\hat{x}_i)_{i=l}^{k},(\xi_i)_{i=l}^{k-1},(\nu_i)_{i=l}^{k} \\ (\acute{\xi}_i)_{i=l}^{k-1},(\acute{\nu}_i)_{i=l}^{k}}} \Gamma(\hat{x}_l) + \sum_{i=l}^{k}(1-\lambda_j)L_i(\xi_i,\nu_i) + \tag{7a}$$

$$\sum_{i=l}^{k}\lambda_j L_i(\acute{\xi}_i,\acute{\nu}_i) \tag{7b}$$

$$\text{s.t.} \quad \hat{x}_{i+1} = f(\hat{x}_i) + \acute{\xi}_i, \qquad i = l,\dots,k-1, \tag{7c}$$

$$\hat{x}_{i+1} = f(0) + F_i\hat{x}_i + \xi_i, \qquad i = l,\dots,k-1, \tag{7d}$$

$$y_i = h(\hat{x}_i) + \acute{\nu}_i, \qquad i = l,\dots,k, \tag{7e}$$

$$y_i = h(0) + H_i\hat{x}_i + \nu_i, \qquad i = l,\dots,k, \tag{7f}$$

$$\big((\hat{x}_i)_{i=l}^{k},(\xi_i)_{i=l}^{k-1}\big) \in \Omega, \tag{7g}$$

$$\big((\hat{x}_i)_{i=l}^{k},(\acute{\xi}_i)_{i=l}^{k-1}\big) \in \Omega, \tag{7h}$$

where the strictly monotonic increasing sequence $\Lambda = \{\lambda_0, \lambda_1, \dots, \lambda_{n_\lambda}\}$ is defined for $0 \leq \lambda_j \leq 1$ and $j \in \{0, 1, \dots, n_\lambda\}$. In the case that $\lambda_0 = 0$, the equality constraints (7c), (7e) and the corresponding disturbance variables can be eliminated, such that the problem becomes convex as long as the inequality constraints are convex. Starting from $\lambda_0 = 0$ ensures that Algorithm 1 ends up with a unique solution, since the optimization problem becomes convex. Although, due to the approximation mentioned in Remark 3, the solution can be different from the global solution of the MHE. Gradually, as λ goes to 1, the optimization problem adds more weight to satisfy the nonlinear dynamics. Finally, for $\lambda = 1$, the constraints (7d), (7f) and the corresponding disturbance variables can be eliminated. It should be noted that the resulting homotopy path is followed within every sampling period of the MHE scheme. Even though it is not guaranteed, by tuning the n_λ homotopy steps in (7) as $\lambda \to 1$, the resulting local minimum of this non-convex problem can be close or equal to the global solution. This might be possible by assuming sufficiently small bounded disturbances, which will be illustrated through two different case studies in Sect. 5.

Remark 5: It can be assumed that bounds on w_i, v_i, $\acute{w}_i$ and $\acute{v}_i$ for $i = l, \dots, k$ are strict enough, such that it guarantees the solution of CMHE to be sufficiently close to the solution of the nonlinear MHE problem. It should be noted that there is generally no guarantee of obtaining the global solution. However, using the proposed method (HMHE), we might expect to have a higher chance of finding a global solution (or maybe a better local solution), compared to the case of applying a gradient based optimization solver directly to the non-convex problem.

The procedure presented in Alg. 1 provides an insight on how to implement the homotopy based MHE scheme. Although this method can further improve the CMHE solution, it comes at the price of solving a sequence of n_λ parameterized optimization problems. Hence, this motivates the study of a computationally attractive way to implement this algorithm. Note that Algorithm 1 assumes a local solver to be used for the nonlinear optimization problem in Step 4, based on the previous solution along the homotopy path as an initial guess.

4.4 *Adaptive tuning of the HMHE*

In some examples, solving the convex optimization problem (CMHE) for $\lambda_0 = 0$ in the first step, results in a solution that is close enough to the global minimum. In this case, no intermediate steps might be needed to solve the nonlinear optimization problem for $\lambda = 1$. Therefore, an adaptive step size selection procedure can be used for the numerical continuation method as discussed in Allgower

Algorithm 1 Implementation of HMHE

Initialization: $\hat{x}_0 \in \mathbb{X}_0$, $\xi_0 \in \mathbb{W}$, $\acute{\xi}_0 \in \mathbb{W}$, $\nu_0 \in \mathbb{V}$, $\acute{\nu}_0 \in \mathbb{V}$, $\lambda_0 = 0$ and $j = 0$
Input: n_λ, $\mathcal{Y}_l^k$
Compute the LTV model (2) using $\mathcal{Y}_l^k$
Formulate HMHE using (1) and (2) as in Eq. (7)
while $j < n_\lambda$ **do**
  Solve HMHE(λ_j) in (7) and update the optimizers: $(\hat{x}_i)_{i=l}^k$, $(\xi_i)_{i=l}^{k-1}$, $(\nu_i)_{i=l}^k$, $(\acute{\xi}_i)_{i=l}^{k-1}$ and $(\acute{\nu}_i)_{i=l}^k$
  Use the solution of HMHE(λ_j) as an initial guess for the next problem
  $j \leftarrow j + 1$
end while
Output: $\hat{x}_k$.

& Georg (1990). Algorithm 2 presents a simplified implementation of such a procedure, by adaptively changing the total number of steps n_λ with a factor of δ_λ. This is equivalent with changing the step size in Λ. In the case that subsequent solutions are sufficiently close, i.e. small $\|\mathcal{X}_{j+1} - \mathcal{X}_j\|$, then the number of homotopy steps can be reduced. On the other hand, if these solutions are relatively different, then the number of steps can also be increased. The details are summarized in Alg. 2, where the small δ_x and large Δ_x are the tuning variables. Simulation results show a significant reduction of the computational complexity, compared to the standard implementation in Algorithm 1.

Algorithm 2 Implementation of modified HMHE

Initialization: $\mathcal{X}_0 = \hat{x}_0 \in \mathbb{X}_0$, $\xi_0 \in \mathbb{W}$, $\acute{\xi}_0 \in \mathbb{W}$, $\nu_0 \in \mathbb{V}$, $\acute{\nu}_0 \in \mathbb{V}$, $\lambda_0 = 0$ and $j = 0$
Input: n_λ, $\mathcal{Y}_l^k$, $0 < \delta_\lambda < 1$, $\delta_x > 0$ and $\Delta_x > 0$
Compute the LTV model (2) using $\mathcal{Y}_l^k$
Formulate HMHE using (1) and (2) as in Eq. (7)
while $j < n_\lambda$ **do**
  Solve HMHE(λ_j) in (7) and update the optimizers: $(\hat{x}_i)_{i=l}^k$, $(\xi_i)_{i=l}^{k-1}$, $(\nu_i)_{i=l}^k$, $(\acute{\xi}_i)_{i=l}^{k-1}$ and $(\acute{\nu}_i)_{i=l}^k$
  $\mathcal{X}_{j+1} \leftarrow (\hat{x}_i)_{i=l}^k$
  if $\|\mathcal{X}_{j+1} - \mathcal{X}_j\| < \delta_x$ **then**
    $n_\lambda \leftarrow \delta_\lambda n_\lambda$
  else if $\|\mathcal{X}_{j+1} - \mathcal{X}_j\| > \Delta_x$ **then**
    $n_\lambda \leftarrow n_\lambda / \delta_\lambda$
  end if
  Use the solution of HMHE(λ_j) as an initial guess for the next problem
  $j \leftarrow j + 1$
end while
Output: $\hat{x}_k$.

Remark 6: In Algorithm 2 the increase/decrease of n_λ might not be well-posed for arbitrary δ_λ. However it is assumed that n_λ stays integer; i.e. the increment/decrement is a multiplier of n_λ.

5. Numerical case studies

This section presents two numerical case studies for state and parameter estimation, in order to illustrate the proposed method. As mentioned before, input and output disturbances (in some literature known as the process and measurement noise) might reduce the validity of the LTV model in (2). Therefore, we will first investigate the influence of zero mean random input and

output disturbances with small amplitude on the convergent properties of HMHE. In the second example, a biased output disturbance is assumed in two forms: constant and random. This type of disturbances ended up with a deviated estimates for CMHE and consequently deteriorated the quality of HMHE estimates. Nevertheless, it will be illustrated that the homotopy approach can result in a better local minimum compared to the MHE.

5.1 *First case study: the effect of small disturbances*

Consider the parameter varying one-dimensional nonlinear system

$$\begin{aligned} x_{k+1} &= (1 - 5T_s + 5T_s x_k)x_k + p_k \cos(x_k) + \acute{w}_{k,1} =: f(x_k, p_k) + \acute{w}_{k,1}, \\ y_k &= x_k + v_k, \end{aligned}$$

where p_k denotes the parameter which is assumed to be slowly time-varying in between each sampling interval $T_s = 0.01$ s. It is worth mentioning that the domains considered in Assumption 3 are compact, hence the functions f and h are Lipschitz continuous in all their arguments on the respective compact domains. Let us consider the compact sets as $-1 \leq x_k \leq 1$ and $-5 \leq p_k \leq -1$. Choosing $N = 2$ and $\phi(\begin{bmatrix} x_k & p_k \end{bmatrix}^{\mathrm{T}}) = c_x\|x_k\|^2 + c_p\|p_k\|^2$ gives the uniform observability based on Def. 2, where c_x and c_p are chosen as sufficiently small positive constants. This property is mainly the result of the compactness sets assumption. Further on this topic, it should be mentioned that compactness of the domain is not only linked to the boundedness of states, but also to the boundedness of estimates. In the latter case constraints on state estimates can be helpful for the theory although may not be needed in practice. This is considered as one of our ongoing research topic.

Moreover, the input and output disturbances are white Gaussian noise with variance 0.001 and 0.005, respectively. In order to formulate this problem in an augmented parameter and state estimation framework, we consider p as a second state and its dynamic reads $p_{k+1} = p_k + \acute{w}_{k,2}$. The LTV model (2) can be obtained as follows

$$\begin{bmatrix} x_{k+1} \\ p_{k+1} \end{bmatrix} = \begin{bmatrix} 1 - 5T_s + 5T_s y_k & \cos y_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ p_k \end{bmatrix} + \begin{bmatrix} w_{k,1} \\ \acute{w}_{k,2} \end{bmatrix} = F_k \begin{bmatrix} x_k \\ p_k \end{bmatrix} + \begin{bmatrix} w_{k,1} \\ \acute{w}_{k,2} \end{bmatrix},$$

where we note that $w_{k,1}$ is not necessarily white and can be denoted as $w_{k,1} = w(y_k, \acute{w}_{k,1}, v_k)$. The cost function for the MHE problem formulation in (5) is given by

$$J_0 = P_x(x_l - \bar{x}_l)^2 + P_p(p_l - \bar{p}_l)^2 + \sum_{i=l}^{k-1} Q_x(x_{i+1} - f(x_i, p_i))^2 + Q_p(p_{i+1} - p_i)^2 + \sum_{i=l}^{k} R_x(y_i - x_i)^2,$$

where $N = 10$ such that $l = k - 9$ and the $2N = 20$ optimization variables are $(x_i)_{i=l}^{k}$ and $(p_i)_{i=l}^{k}$. The weights are selected as $P_x = P_p = Q_p = 1$, $Q_x = 1000$ and $R_x = 200$. Assuming that the system is in steady state, for different values of $(x_j)_{j=l}^{k} := x$ and $(p_j)_{j=l}^{k} := p$, the non-convex cost function of the MHE is illustrated in Fig. 3. The global solution is denoted by a white circle.

The Hessian of this cost function (HMHE) is denoted by $H \in \mathbb{R}^{2N \times 2N}$. Based on the eigenspectrum of the corresponding Hessian matrix, in Fig. 4(a), the $2N$ eigenvalues of the Hessian for the HMHE problem are illustrated, as λ is changing from 0 to 1. The optimizer index denotes each element in the augmented state and parameter space over the prediction horizon. To investigate the convexity of the HMHE problem, the state and parameter estimates used in the Hessian and its eigenvalues corresponding to each optimizer indices, are plotted for one single time instance in Fig. 4(a). For $\lambda = 0$, the problem is convex, such that all the eigenvalues are positive. As the value of the homotopy parameter changes and gets closer to 1, we can observe that some eigenvalues are

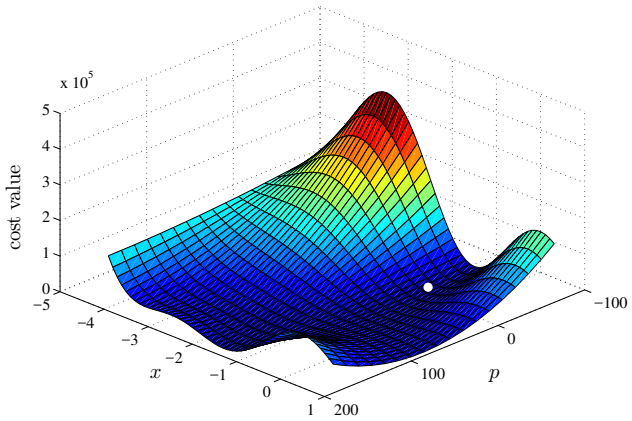
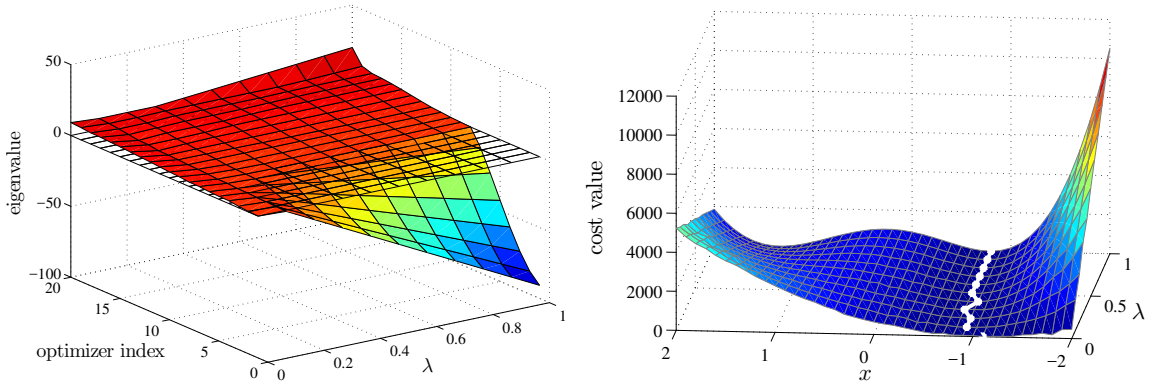


Figure 3. Convexity analysis of MHE for the first numerical case study (white circle is the global solution with no disturbances included).



(a) Hessian eigenvalues of the HMHE.

(b) Cost function of HMHE with the homotopy path (white line).

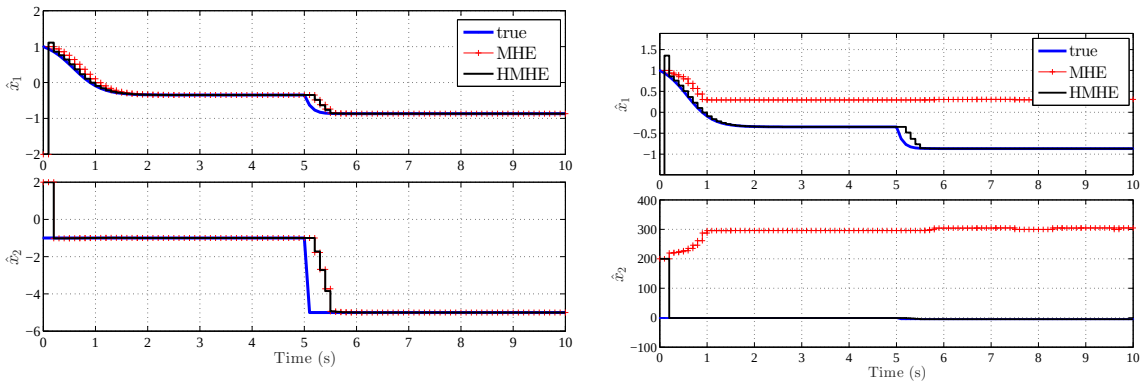
Figure 4. Convexity analysis of HMHE for the first numerical case study.

negative. In Fig. 4(b), the cost function of HMHE is illustrated for different λ, where the effect of disturbances is visible, even though it is small. For this illustration, it is assumed that the parameter is fixed and the system is in steady state, i.e., the value of the augmented state is constant over the horizon. In this case, we can demonstrate the behavior of HMHE for different values of x. The homotopy path is denoted by the white curve and it shows the transition between the result of the convex problem and the non-convex one.

Performance analysis

The simulations were carried out on an Ubuntu 16.04 desktop computer using the ACADO Toolkit (Houska et al., 2011). Table 1 shows that the performance of different algorithm implementations for the proposed homotopy based HMHE scheme result in a similar level of accuracy. Denoting the augmented estimated state as $\hat{x} = \begin{bmatrix}\hat{x}_1 & \hat{x}_2\end{bmatrix}$, with a proper initialization $\hat{x} = \begin{bmatrix}-2 & 2\end{bmatrix}$ close to the true value $\begin{bmatrix}x_0 & p_0\end{bmatrix} = \begin{bmatrix}1 & -1\end{bmatrix}$, all the methods have indistinguishable estimation errors; see Fig. 5(a). The rest of 20 optimization variables are updates with the system update equation for $k < N$.

The non-convex MHE formulation allows a local solver to find an unacceptable local solution, given a bad initial guess; c.f. Fig. 5(b). In this scenario, the initial state is assumed to be far from the true value as it was chosen $\hat{x} = \begin{bmatrix}0 & 200\end{bmatrix}$. The HMHE approach converged to the desired solution, while MHE converged to one of the alternative local minima. A comparison study in the sense of sum of absolute estimation error is provided in Table 1. It should be noted that considering



(a) Estimation results with true initialization.

(b) Estimation results with wrong initialization.

Figure 5. HMHE simulation results for the first numerical case study.

the number of optimizers, which in this case is 20, it is not feasible to demonstrate all the local minima. For this numerical case study, the computational complexity of HMHE using Algorithm 2 is relatively faster compared with the nominal MHE; see Table 1. Further on the computational complexity, it should be noted that the paper assumes that the mapping ψ is readily available hence its calculation does not add to the overall computational complexity; see Abdollahpouri et al. (2017, Sect. 5.2) for an example with 4 states. For more complicated examples the solution of integrals might have to be computed at each time step to compute the matrices F and H. The numerical complexity of computing F and H, if ψ is available, is generally low compared to solving the HMHE problem.

Table 1. Comparison of average computation time and overall estimation error for the first case study.

	MHE	Alg. 1	Alg. 2
Estimation error (wrong x_0)	1850	10.45	10.45
Estimation error (true x_0)	5.26	4.03	4.03
Execution time (ms)	0.63	0.7	0.3

5.2 *Second case study: the effect of a large output disturbance*

In this case study, a primitive gradient descent method has been adopted to demonstrate the possibility of finding a local minimum with the HMHE approach. It should be noted that a different solver might result in different estimation performance. Let us consider the following scalar nonlinear dynamics

$$x_{k+1} = T_s(-20x_k^3 + 10x_k^2 + u_k) + x_k,$$
$$y_k = x_k + v_k,$$

where v_k is primarily assumed to be a fixed output disturbance (measurement bias), which is 1.5 in this example. The input signal is assumed to be a constant $u_k = 30$ in the beginning of the test and has been changed to another value $u_k = -15$ in the middle of the simulation. The sampling period T_s is 0.01 s and the true initial state is $x_0 = -1$. The LTV model can be obtained, utilizing the measurement y_k as $x_{k+1} = (-20T_s y_k^2 + 10T_s y_k + 1)x_k + T_s u_k + w_k$. In this transformation, the influence of v_k is summarized by $w_k : w(y_k, v_k)$. Similar to the previous example, on a compact domain, the nonlinear function f is Lipschitz, since it is continuously differentiable with respect to x. Furthermore, assuming $N = 1$ and ϕ as identity function, this nonlinear system is uniformly

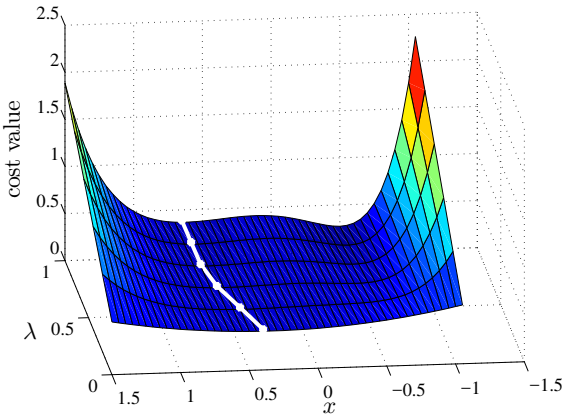
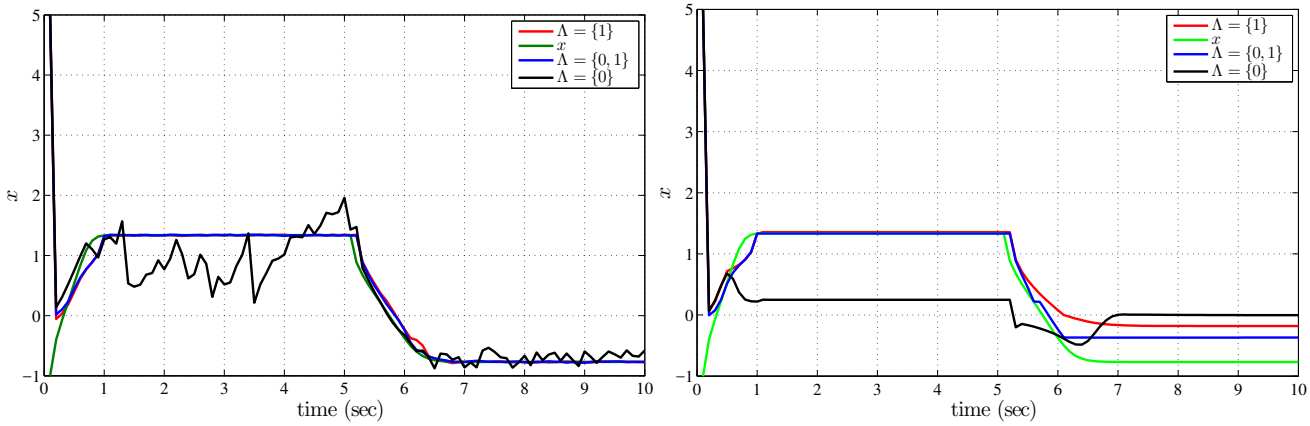


Figure 6. HMHE cost for different x and λ (white line is the homotopy path).



(a) Estimated state with random output disturbance.

(b) Estimated state for constant output disturbance.

Figure 7. HMHE analysis for the second numerical case study.

observable based on the Def. 2. Assuming a horizon $N = 10$ and the cost function in (7), the homotopy transformation is illustrated in Fig. 6 in steady state. The cost function can be simplified as

$$J_2 = P_x(x_l - \bar{x}_l)^2 + \sum_{i=l}^{k-1} Q_x(x_{i+1} - f(x_i, p_i))^2 + \sum_{i=l}^{k} R_x(y_i - x_i)^2,$$

where $P_x = Q_x = R_x = 1$. The white curve shows the homotopy path, which drives the solution to a local minimum. In Fig. 7(b), $\Lambda = \{0\}$ and $\Lambda = \{1\}$ denote the CMHE and MHE, respectively. For the homotopy based case $\Lambda = \{0, 1\}$, the solution got closer to the true value and this can be considered as a better local minimum; see Table 2 for a comparison of the estimation performance.

Furthermore, to demonstrate the benefit of HMHE with at least $n_\lambda = 2$, a white Gaussian noise with variance of 0.5 is assumed for v_k. In this case, even though the CMHE solution deviates at some iterations (see Fig. 7(a)), the homotopy approach can drive the solution to the global minimum. For this case, the tuning parameters in CMHE will stays as before, however the tuning variables in the MHE cost is chosen as $R_x = 0.001$ to reduce the effect of noisy measurements. Therefore, the state update that the MHE is using has more impact than the one in measurement error ($Q_x > R_x$) and eventually helps the HMHE approach to converge to the global minimum. In Table 2, the improvement of performance in the sense of the sum of absolute value of estimation error is illustrated.

Table 2. Comparison of overall estimation error for the second case study

Estimation Method	Biased disturbance	Random disturbance
CMHE	86.96	31.31
MHE	35.42	10.64
HMHE	24.61	10.1

6. Conclusions

This paper presents a homotopy based nonlinear MHE method, which is globally convergent in a nominal condition; with no input and output disturbances. A model reformulation has been employed to transform the nonlinear dynamics into an LTV model, hence the leading optimization problem is convexified—CMHE. This model transformation is sensitive to input and output disturbances, therefore the resulting CMHE problem might have a deviated solution compared to the original nonlinear problem. Although the model transformation is applicable to difference flat systems, this approach is not limited to this class of nonlinear systems. Utilizing a homotopy approach starting with the CMHE solution can lead the overall optimization problem to be convergent, assuming small bounded disturbances. On the other hand, a larger input or output disturbance might deviate the CMHE solution from the global minimum and result in a sub-optimal solution as it is illustrated through two simulation studies.

Acknowledgment

The authors gratefully acknowledge the contribution of the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 607957 (TEMPO); the financial contribution of the STU Grant Scheme for the Support of Excellent Teams of Young Researchers, the Slovak Research and Development Agency (APVV) under the contract APVV-14-0399; the support of the Scientific Grant Agency (VEGA) of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the contract 1/0144/15; the Research Council of Norway, Statoil, DNV GL and Sintef through the Centers of Excellence funding scheme, project number 223254 - Centre for Autonomous Marine Operations and Systems (NTNU-AMOS); and grant 250275; and the support of EEA Grant no. SK06-II-01-004; Support by the EU via ERC-HIGHWIND (259166) and ITN-AWESCO (642682), by DFG via Research Unit FOR 2401; support by the German BMWi via project eco4wind, as well as the Belgian FWO via a PhD grant for R.Q. is gratefully acknowledged.

References

Abdollahpouri, M., Gulan, M., Takács, G., et al. (2016). Real-time state estimation for an adaptive vibration energy harvesting system. In *14th IFAC conference on programmable devices and embedded systems PDES 2016* (Vol. 49, pp. 127–132).

Abdollahpouri, M., Haring, M., Johansen, T. A., Takács, G., & Rohaľ-Ilkiv, B. (2017). Nonlinear state and parameter estimation using discrete-time double Kalman filter. In *The 20th world congress of the IFAC.* (Accepted, online: http://bit.ly/2jC5sbW)

Abdollahpouri, M., Takács, G., & Rohal'-Ilkiv, B. (2017). Real-time moving horizon estimation for a vibrating active cantilever. *Mechanical Systems and Signal Processing*, *86*, 1–15.

Abraham, R., Marsden, J. E., & Ratiu, T. (2012). *Manifolds, tensor analysis, and applications.* Springer.

Allgower, E. L., & Georg, K. (1990). *Introduction to numerical continuation methods.* Colorado State University Press.

Bonilla, J., Diehl, M., Logist, F., De Moor, B., & Van Impe, J. (2010a). An automatic initialization procedure in parameter estimation problems with parameter-affine dynamic models. *Computers & chemical engineering*, *34*(6), 953–964.

Bonilla, J., Diehl, M., Logist, F., De Moor, B., & Van Impe, J. F. (2010b). A convexity-based homotopy method for nonlinear optimization in model predictive control. *Optimal Control Applications and Methods*, *31*(5), 393–414.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Diehl, M., Ferreau, H. J., & Haverbeke, N. (2009). *Efficient numerical methods for nonlinear MPC and moving horizon estimation* (L. Magni, D. M. Raimondo, & F. Allgöwer, Eds.) (No. 384). Springer-Verlag.

Dinh, Q. T., Savorgnan, C., & Diehl, M. (2011). Real-time sequential convex programming for nonlinear model predictive control and application to a hydro-power plant. In *2011 50th ieee conference on decision and control and european control conference* (pp. 5905–5910).

Fliess, M., Lévine, J., Martin, P., & Rouchon, P. (1995). Flatness and defect of non-linear systems: introductory theory and examples. *International journal of control*, *61*(6), 1327–1361.

Haseltine, E. L., & Rawlings, J. B. (2005). Critical evaluation of extended Kalman filtering and moving-horizon estimation. *Industrial & engineering chemistry research*, *44*(8), 2451–2460.

Houska, B., Ferreau, H. J., & Diehl, M. (2011). ACADO toolkit - An open-source framework for automatic control and dynamic optimization. *Optimal Control Applications and Methods*, *32*(3), 298–312.

Hovgaard, T. G., Boyd, S., Larsen, L. F., & Jørgensen, J. B. (2013). Nonconvex model predictive control for commercial refrigeration. *International Journal of Control*, *86*(8), 1349–1366.

Johansen, T. A. (2011). Introduction to nonlinear model predictive control and moving horizon estimation. *Selected Topics on Constrained and Nonlinear Control*(1), 1–53.

Johansen, T. A., & Fossen, T. I. (2016a). The eXogenous Kalman filter (XKF). *International Journal of Control*, 1–7.

Johansen, T. A., & Fossen, T. I. (2016b). Nonlinear filtering with eXogenous Kalman filter and double Kalman filter. In *European control conference* (pp. 1722–1727).

Kühl, P., Diehl, M., Kraus, T., Schlöder, J. P., & Bock, H. G. (2011). A real-time algorithm for moving horizon state and parameter estimation. *Computers & Chemical Engineering*, *35*, 71–83.

Liu, J. (2013). Moving horizon state estimation for nonlinear systems with bounded uncertainties. *Chemical Engineering Science*, *93*, 376 - 386.

Rao, C. V., Rawlings, J. B., & Mayne, D. Q. (2003). Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, *48*, 246–258.

Sira-Ramirez, H., & Agrawal, S. K. (2004). *Differentially flat systems*. CRC Press.

Sira-Ramirez, H., & Castro-Linares, R. (2000). Sliding mode rest-to-rest stabilization and trajectory tracking for a discretized flexible joint manipulator. *Dynamics and Control*, *10*(1), 87–105.